

PENGEKSTRAKAN MAKLUMAT BERDASARKAN PENGECEMAN ENTITI NAMA DARIPADA TEKS TERJEMAHAN AL-QURAN DALAM BAHASA INGGERIS

Shasha Arzila binti Tarmizi

Dr. Saidah binti Saad

ABSTRAK

Al-Quran merupakan kitab suci bagi umat Islam dan mengandungi pelbagai maklumat yang dijadikan panduan dalam kehidupan seorang Muslim. Pengetahuan yang terkandung dalam Al-Quran antaranya ialah kisah para Nabi, hukum dalam berjual beli (muamalat), hukum dalam pernikahan, hukum jenayah dan banyak lagi. Proses pengekstrakan maklumat secara manual akan mengambil masa dan tenaga kerja yang banyak kerana Al-Quran mengandungi lebih daripada enam ribu ayat. Bidang ontologi dalam domain Al-Quran juga masih belum menyeluruh dan masih banyak topik dan entiti yang boleh digunakan sebagai tajuk kajian dalam bidang ontologi ini. Projek ini dijalankan bertujuan untuk mengekstrak maklumat yang terkandung di dalam Al-Quran menggunakan teks terjemahan dalam bahasa Inggeris bagi memudahkan pencarian maklumat. Pengekstrakan maklumat ini dijalankan menggunakan tiga teknik pengecaman entiti nama. Tiga teknik yang digunakan dalam projek ini ialah teknik berdasarkan peraturan, teknik menggunakan pustaka sedia ada dan teknik pembelajaran mesin tak berarah. Pustaka yang digunakan dalam ketiga-tiga teknik ini ialah daripada pustaka sumber terbuka SpaCy. Ketiga-tiga teknik ini kemudiannya dibandingkan menggunakan kaedah kejituan, dapatan semula dan skor-F. Projek ini akan dijalankan dalam empat fasa iaitu fasa keperluan, fasa reka bentuk, fasa pelaksanaan dan akhir sekali ialah fasa penilaian. Terdapat penambahan entiti baru yang dibina untuk disesuaikan bagi domain Al-Quran seperti nama Nabi, nama-nama Allah dan nama golongan. Penilaian yang dijalankan menunjukkan teknik berasaskan peraturan mempunyai keputusan yang lebih baik daripada dua teknik yang lain. Projek ini dapat memberi panduan dalam proses pengekstrakan maklumat daripada Al-Quran menggunakan komponen pemprosesan bahasa tabii. Hasil pengekstrakan maklumat daripada projek ini juga dapat digunakan dalam bidang ontologi Al-Quran yang sedang berkembang.

1 PENGENALAN

Al-Quran merupakan kitab suci bagi penganut agama Islam. Kitab yang sudah berusia ribuan tahun ini mengandungi banyak pengetahuan dan informasi yang dijadikan sebagai panduan hidup bagi umat Islam. Pelbagai kategori maklumat dapat dijumpai di dalam Al-Quran antaranya bab akidah, kisah para Nabi, kisah umat pada zaman

dahulu, hukum-hukum jenayah, bab-bab perkahwinan, urusan jual beli dan banyak lagi. Maklumat-maklumat ini tidak disusun mengikut tajuk di dalam Al-Quran. Satu surah boleh mengandungi bermacam-macam jenis maklumat dan kedudukan maklumat-maklumat ini tidak tersusun maka agak sukar sekiranya seseorang perlu menelusuri kesemua 6236 ayat yang terkandung di dalam Al-Quran untuk mendapatkan maklumat secara manual.

Pengecaman entiti nama atau dalam bahasa Inggeris disebut sebagai *Named Entity Recognition* (NER), merupakan salah satu komponen dalam pemrosesan bahasa tabii. Komponen ini juga merupakan salah satu proses penting dalam pengekstrakan maklumat daripada data berbentuk teks yang tidak berstruktur dalam bahasa tabii. Hal ini kerana entiti nama dapat membantu pengguna dengan mudah mengenal pasti topik utama sesuatu dokumen. Pengguna dapat mengenal pasti elemen penting dalam teks, seperti nama orang, tempat, nilai wang, dan banyak lagi. Mengekstrak entiti utama dalam teks bukan sahaja dapat membantu menyusun data tidak berstruktur dan mengesan maklumat penting, malahan proses ini sangat penting jika pengguna harus berurusan dengan set data teks yang besar.

Projek ini akan menggunakan komponen NER dalam mendapatkan maklumat daripada teks terjemahan Al-Quran dalam bahasa Inggeris. Dataset yang digunakan dalam projek ini ialah teks terjemahan Al-Quran dalam bahasa Inggeris oleh Dr. Mohammad Taqi-ud-Din Al-Hilali dan Dr. Muhammad Muhsin Khan. Terjemahan dalam bahasa Inggeris digunakan kerana kebanyakan pustaka bagi NER telah dibina untuk teks bahasa Inggeris terutama pustaka SpaCy yang digunakan dalam projek ini.

2 PENYATAAN MASALAH

Al-Quran mengandungi 114 surah dan surah-surah panjang biasanya mengandungi pelbagai jenis maklumat yang berlainan. Sebagai contoh, cerita berkenaan seseorang Rasul ada yang terkandung dalam beberapa surah yang berbeza. Seseorang individu akan mengalami kesukaran untuk mendapatkan rujukan berkenaan satu-satu maklumat kerana faktor ini. Proses mengklasifikasi dan mengekstrak maklumat secara manual akan mengambil masa dan tenaga kerja yang banyak dan cenderung kepada kesilapan

atau kecuiaan. Selain itu, zaman moden ini sedang menunjukkan perkembangan dalam bidang teknologi semantik seperti ontologi Al-Quran. Teknologi semantik adalah satu bidang yang membolehkan komputer memahami makna perkataan dan membantu manusia dalam memproses perkataan yang terdapat pada suatu domain. Ontologi Al-Quran yang sedia ada juga masih belum menyeluruh dan masih terhad kepada suatu topik (Maha Al-Yahya et al. 2010; Azman Taa et al. 2014) atau surah tertentu (Rohana Ismail et al. 2015). Perkembangan ontologi Al-Quran masih tertumpu kepada topik-topik yang berkaitan akidah walhal masih banyak perkara atau elemen yang boleh dijadikan bahan kajian seperti nama-nama tempat yang disebut di dalam Al-Quran ataupun nama-nama tokoh. Elemen-elemen ini boleh diekstrak melalui proses pengecaman nama dan sekali gus membantu memberi lebih banyak idea atau buah fikiran untuk perkembangan ontologi Al-Quran.

Tambahan pula, terdapat pelbagai jenis teknik yang boleh digunakan untuk mengekstrak maklumat namun masih belum ada kajian yang menilai kesesuaian suatu teknik pengecaman entiti nama untuk mengekstrak maklumat daripada kitab suci Al-Quran. Al-Quran tidak seperti dokumen atau teks-teks lain kerana kaedah biasa tidak akan dapat mengekstrak maklumat yang terkandung dalam Al-Quran kerana terdapat pelbagai entiti khas yang tidak dapat dijumpai dalam dokumen atau artikel biasa. Justeru, projek ini akan membandingkan tiga teknik iaitu teknik berpanduan peraturan, teknik menggunakan pustaka asas yang sedia ada dan teknik yang menggunakan pembelajaran mesin berarah untuk menentukan kesesuaian dalam menguruskan teks domain seperti Al-Quran.

3 OBJEKTIF KAJIAN

Projek ini bertujuan untuk:

1. Melakukan pengecaman entiti nama yang terkandung di dalam Al-Quran.
2. Mengekstrak segala maklumat berasaskan pengecaman entiti nama.
3. Melakukan perbandingan antara tiga teknik yang digunakan iaitu teknik berasaskan peraturan, teknik menggunakan pustaka asas dan teknik pembelajaran mesin berarah.
4. Menentukan teknik terbaik berdasarkan perbandingan yang telah dibuat.

4 METOD KAJIAN

Model proses pembangunan atau metodologi yang akan digunakan dalam projek ini ialah model air terjun (waterfall). Model ini sesuai bagi projek yang mementingkan perancangan dan struktur (Wilfred Van Casteren 2017). Setiap fasa perlu dilengkapkan sebelum ke fasa seterusnya seperti yang diperlukan oleh projek ini. Projek tidak dapat dijalankan sekiranya proses pembinaan model latihan tidak lengkap.

4.1 Fasa Keperluan

Fasa keperluan ini dimulakan dengan proses mengenal pasti dataset yang akan digunakan dalam kajian ini. Dataset yang digunakan dalam penyelidikan ini mempunyai tafsiran yang terperinci bagi setiap istilah yang digunakan dan ia merujuk kepada pentafsir terkemuka iaitu At-Tabari, Al-Qurtubi dan Ibn Kathir (Gabriel Reynolds 2012). Dataset bagi fasa ini iaitu surah Yunus, terlebih dahulu dipra-proses atau dibersihkan demi kemudahan proses NER. Sebelum dibersihkan, kesemua ayat daripada surah Yunus telah disimpan dalam satu fail berformat “.txt”. Kesemua proses ini dijalankan menggunakan bahasa pengaturcaraan Python. Modul penting dalam proses ini ialah ungkapan nalar (RegEx) kerana ungkapan nalar ini digunakan untuk sebagai corak untuk elemen yang perlu dibersihkan.

Pertama sekali, proses pembersihan yang dilakukan ialah pembuangan nombor atau angka. Angka-angka ini tidak mempunyai sebarang kepentingan kerana kebanyakan angka datang daripada penandaan nombor ayat. Tambahan pula, sebarang jumlah atau bilangan yang dinyatakan dalam teks terjemahan tersebut ditulis dalam bentuk perkataan dan bukan angka.

Seterusnya, beberapa tanda baca turut dibuang daripada data iaitu tanda noktah (.), tanda koma (,), tanda seru (!), tanda tanya (?), titik bertindih (;) dan pembuka atau penutup kata (“”). Kesemua tanda baca ini tidak mempunyai sebarang signifikan dalam proses pengecaman. Namun begitu, simbol kurungan tidak dibuang kerana

ianya akan digunakan untuk proses pra-pemrosesan yang lain. Proses pembersihan ini juga dilakukan menggunakan ungkapan nalar.

Setiap surah dalam teks terjemahan dimulakan dengan pengenalan surah yang mengandungi beberapa maklumat berkenaan surah tersebut. Maklumat-maklumat ini dikeluarkan daripada dataset menggunakan ungkapan nalar kerana ia tidak mengandungi sebarang entiti yang signifikan untuk diekstrak. Elemen lain yang disingkirkan daripada teks ialah ayat yang ditulis dalam kurungan. Setelah semua langkah pra-pemrosesan dijalankan, dataset kini hanya dipenuhi dengan terjemahan yang betul-betul disebut dalam Al-Quran tanpa ada data-data yang tidak relevan. Langkah pra-pemrosesan asas yang lain seperti tokenisasi tidak dijalankan kerana langkah ini telah tersedia dalam pustaka SpaCy yang akan digunakan dalam proses berikutnya. Pembuangan huruf besar juga tidak dilaksanakan kerana peranan huruf besar penting dalam pelaksanaan NER yang mengancam kebanyakan entiti yang bermula dengan huruf besar.

4.2 Fasa Reka Bentuk

Terdapat tiga jenis teknik yang diguna pakai dalam kajian ini iaitu teknik berasaskan peraturan, teknik menggunakan pustaka sedia ada dan teknik pembelajaran mesin berarah. Peraturan yang digunakan dalam teknik pertama akan dibina dalam fasa ini. Teknik kedua pula tidak memerlukan apa-apa persediaan atau reka bentuk kerana teknik ini akan menggunakan pustaka SpaCy yang sedia ada. Model NER bagi teknik pembelajaran mesin turut direka bentuk dalam fasa ini. Hanya sebahagian dataset akan digunakan sebagai model latihan dan sebahagian lagi akan digunakan dalam pengujian. Dataset terdiri daripada 70 ayat Al-Quran yang telah dilabelkan dengan entiti nama yang berkaitan.

a. Teknik berasaskan peraturan

Teknik ini memerlukan pembinaan peraturan yang akan digunakan dalam fasa ketiga. Kategori entiti seperti nama Nabi, nama Allah, istilah khas masa dan nama golongan

dibina menggunakan peraturan dalam fasa ini. Peraturan bagi teknik ini dibina menggunakan ungkapan nalar dengan bantuan enjin pepadanan peraturan SpaCy.

b. Teknik menggunakan pustaka sedia ada

Pustaka sedia ada bagi teknik ini ialah pustaka sumber terbuka SpaCy. SpaCy ialah pustaka bagi pemprosesan bahasa tabii. Terdapat pelbagai tugas yang boleh dilakukan menggunakan SpaCy seperti tokenisasi, lemmatisasi termasuklah pengecaman entiti (Neelam Tyagi, 2020). Teknik ini digunakan dalam projek ini sebagai perbandingan kepada keperluan pengkhususan entiti dalam menguruskan teks domain Al-Quran kerana model pengecaman entiti nama ini akan mengekstrak asas seperti nama, lokasi dan angka. Teknik ini tidak memerlukan latihan model kerana komponen NER SpaCy telah dipra-latih.

c. Teknik pembelajaran mesin berarah

Teknik ini tidak memerlukan pembinaan peraturan sebaliknya memerlukan pembinaan model latihan. Model latihan dibina menggunakan set latihan yang mengandungi 70 ayat daripada Al-Quran yang telah dilabelkan dengan entiti nama. Projek ini akan menggunakan talian paip SpaCy untuk melatih model. SpaCy mempunyai talian paip yang boleh diubahsuai dan dilatih. Terdapat beberapa komponen yang terdapat dalam talian paip tersebut seperti penghurai dan pengecam entiti nama. Model pembelajaran mesin dalam (deep learning) yang diberi nama 'thinc' telah dipasang dan diintegrasikan secara dalaman dalam talian paip pustaka tersebut. Situasi ini bermaksud pengguna boleh mencipta dan melatih model pengecaman entiti nama yang khusus menggunakan talian paip SpaCy. Talian paip ini membolehkan pengguna melatih model NER SpaCy untuk mengecam kategori entiti yang baharu. Entiti baharu yang diekstrak ialah nama Nabi, nama Allah, istilah khas masa, nama golongan dan entiti istimewa. Dalam fasa ini, dataset latihan yang mengandungi 70 ayat Al-Quran akan disediakan dan model latihan akan direka bentuk dan seterusnya model akan dilatih menggunakan dataset latihan yang telah disediakan.

4.3 Fasa Pelaksanaan

Fasa ketiga ini merupakan fasa yang penting kerana peraturan dan model latihan yang telah dibina pada fasa sebelumnya akan diaplikasikan pada fasa ini. Fasa pelaksanaan ini akan menggunakan dataset yang sebenar berbanding hanya sebahagian dataset pada fasa sebelumnya. Hasil yang lengkap akan diperolehi dalam fasa ini. Proses pelaksanaan ini akan dijalankan mengikut tiga teknik yang dipilih. Hasil keseluruhan iaitu entiti nama yang terkandung dalam dataset sebenar akan diperolehi. Entiti yang telah diekstrak kemudiannya akan disusun sebelum dibawa kepada fasa penilaian.

4.4 Fasa Penilaian

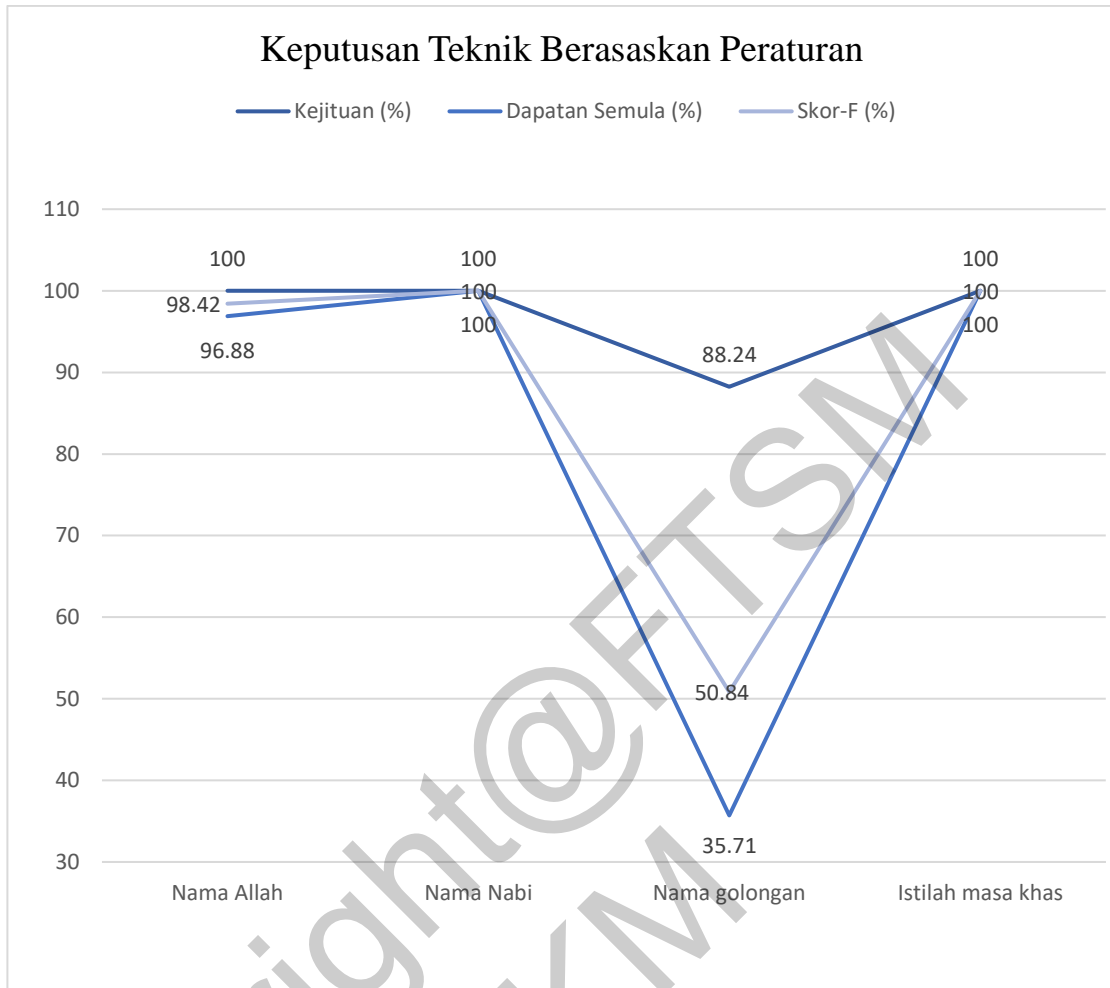
Hasil daripada ketiga-tiga teknik akan disahkan dari segi ketepatan bagi domain Al-Quran. Penilaian ini akan dibuat oleh bantuan pakar yang mahir dengan Al-Quran. Pakar domain yang dipilih ialah tiga orang pelajar daripada Fakulti Pengajian Islam, Universiti Kebangsaan Malaysia. Ketiga-tiga pelajar ini merupakan pelajar dalam jurusan pengajian Al-Quran dan As-Sunnah yang mempunyai latar belakang pendidikan agama dan kepakaran dalam isi kandungan Al-Quran. Ketiga-tiga penilai akan menilai ketiga-tiga teknik menurut pengetahuan yang mereka miliki. Kemudian, berdasarkan penilaian yang telah diperolehi, prestasi setiap teknik akan direkodkan menggunakan nilai kejituan, dapatan semula dan skor-F.

5 HASIL KAJIAN

Keputusan penilaian bagi semua teknik direkodkan dalam jadual berdasarkan keputusan kejituan, dapatan semula dan skor-F masing-masing (Cyril Goutte & Eric Gaussier 2005). Keputusan ini diperolehi setelah penilaian pakar domain dijalankan. Sebelum perbandingan bagi semua teknik, keputusan bagi setiap entiti nama yang diekstrak turut direkodkan dan divisualisasikan dalam bentuk graf.

Teknik berasaskan peraturan

Rajah 5.1 menunjukkan keputusan bagi setiap entiti nama yang berjaya diekstrak oleh teknik ini.

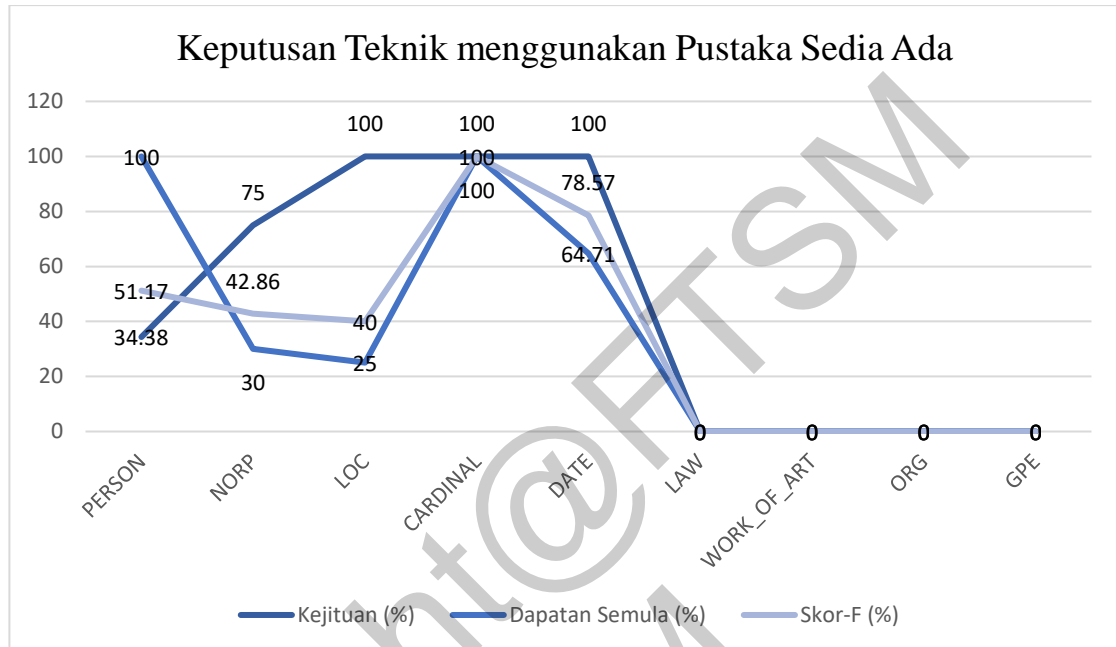


Rajah 5.1: Keputusan teknik berasaskan peraturan

Berdasarkan jadual, gazetir yang digunakan untuk mengekstrak entiti nama Nabi berjaya berfungsi dengan baik. Hal ini demikian kerana gazetir dibina dengan mengambil kira perbezaan dalam ejaan nama-nama Nabi kerana teks terjemahan menggunakan ejaan berbeza bagi sesetengah nama Nabi. Keputusan bagi istilah khas masa juga sangat tinggi kerana kewujudan perkataan 'Day' yang memudahkan pengecaman bagi entiti nama ini. Setelah penilaian oleh pakar domain dilakukan, entiti nama golongan mempunyai nilai dapatan semula yang rendah kerana terdapat banyak entiti yang boleh dikategorikan sebagai golongan namun tidak diekstrak menyebabkan bilangan negatif palsu yang tinggi. Hal ini demikian kerana terdapat perbezaan dalam konteks bahasa Arab dan terjemahan bahasa Inggeris. Purata kejituan teknik ini ialah 97.06 peratus, 83.15 peratus purata dapatan semula dan 87.32 peratus purata skor-F.

Teknik menggunakan pustaka sedia ada

Rajah 5.2 menunjukkan keputusan bagi setiap entiti nama yang berjaya diekstrak oleh teknik ini.

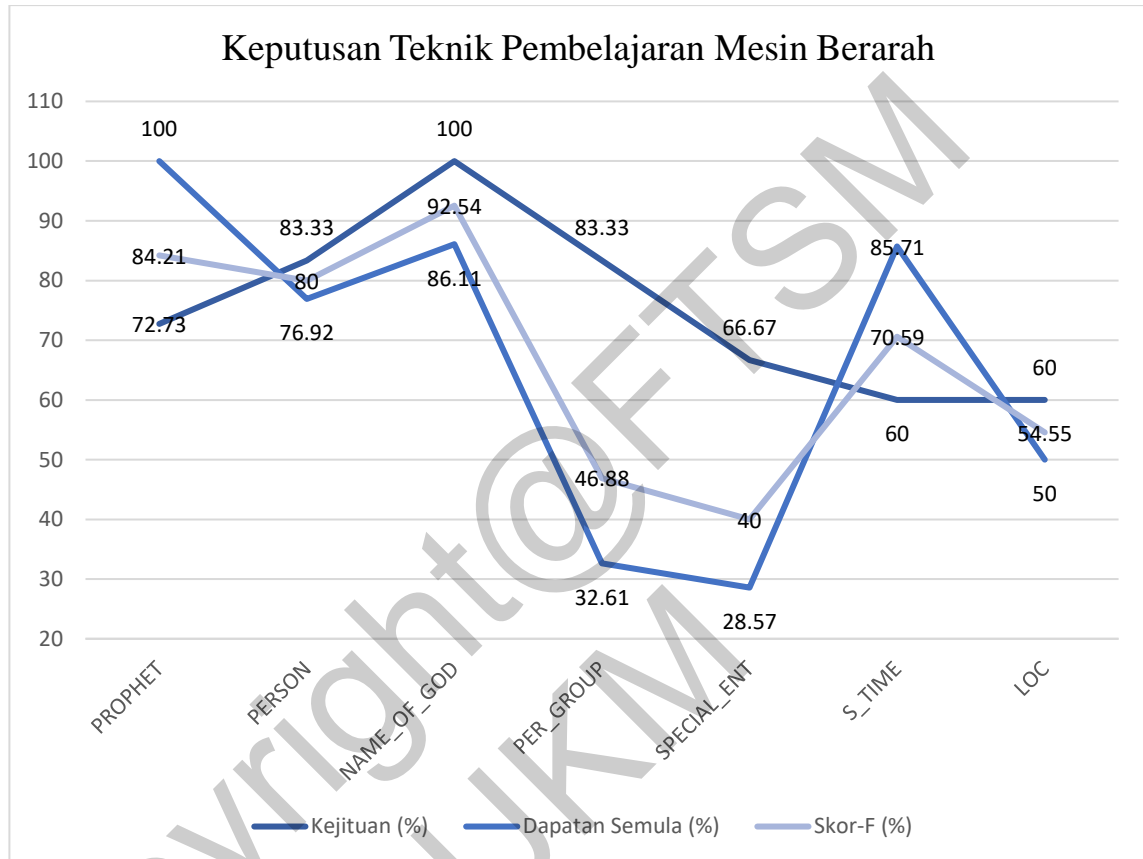


Rajah 5.2: Keputusan teknik menggunakan pustaka sedia ada

Teknik yang menggunakan model NER yang telah dipra-latih oleh SpaCy ini mempunyai pencapaian yang kurang baik dalam mengekstrak entiti seperti ORG, LAW dan WORK_OF_ART. Hal ini berkemungkinan tinggi kerana entiti-entiti ini tidak ditemukan dalam Al-Quran. Dataset Al-Quran yang digunakan mempunyai beberapa entiti GPE seperti Madyan namun model NER ini tidak dapat mengecam entiti tersebut menyebabkan keputusan yang rendah bagi entiti tersebut. Teknik ini hanya mempunyai prestasi yang baik dalam mengekstrak entiti umum seperti tarikh dan kardinal yang tidak khusus kepada mana-mana domain. Namun begitu, teknik ini mempunyai nilai dapatan semula yang baik kerana walaupun teknik ini tidak dapat mengekstrak entiti orang dengan tepat, namun teknik NER ini tidak mengecam entiti orang sebagai entiti lain. Teknik ini memperolehi 45.49 peratus purata kejituan, 35.52 peratus purata bagi semua nilai dapatan semula dan 34.73 peratus bagi purata skor-F.

Teknik pembelajaran mesin berarah

Rajah 5.3 menunjukkan keputusan bagi setiap entiti nama yang berjaya diekstrak oleh teknik ini.



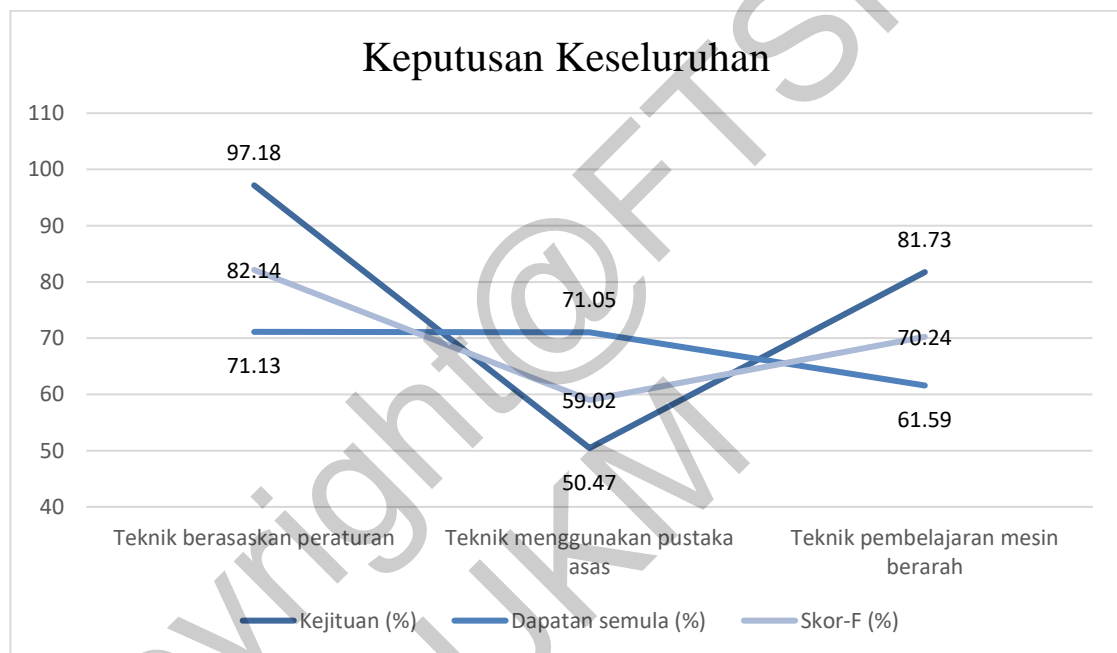
Rajah 5.3: Keputusan teknik pembelajaran mesin berarah

Keputusan penilaian bagi teknik ini menunjukkan yang teknik ini mempunyai kebolehan yang sangat tinggi dalam mengekstrak entiti nama Allah. Pengekstrakan entiti bagi nama golongan (PER_GROUP) mempunyai keputusan dapatan semula yang rendah disebabkan oleh alasan yang sama dalam teknik berasaskan peraturan. Selain daripada entiti nama golongan, entiti nama istilah khas (SPECIAL_ENT) dan entiti tempat (LOC), keputusan bagi entiti lain adalah memuaskan. Carta keputusan juga menunjukkan yang teknik ini berupaya mengecam nama orang dan nama Nabi walaupun kedua-dua entiti mempunyai persamaan iaitu merujuk kepada nama individu. Berdasarkan keputusan dapatan semula, teknik ini mempunyai nilai dapatan semula paling tinggi bagi entiti nama Nabi yang membawa maksud entiti ini tidak

mengecam entiti yang salah sebagai nama Nabi. Purata kejituan bagi teknik ini ialah 75.15 peratus manakala purata dapatan semula ialah sebanyak 65.70 peratus dan purata skor-F sebanyak 66.97 peratus.

Analisis keseluruhan

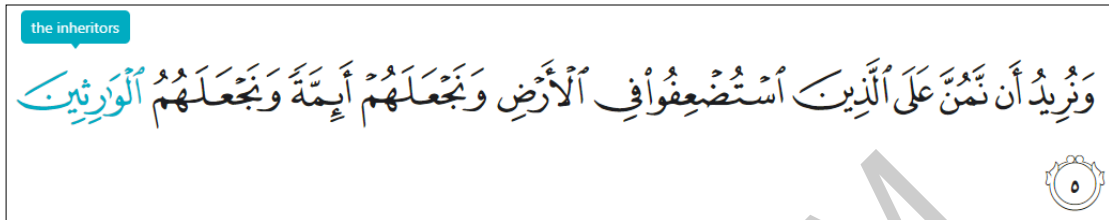
Rajah 5.4 menunjukkan keputusan keseluruhan bagi ketiga-tiga teknik yang telah dinilai oleh pakar domain.



Rajah 5.4: Keputusan keseluruhan

Secara keseluruhan, teknik berasaskan peraturan mempunyai keputusan penilaian terbaik dalam mengekstrak entiti nama daripada Al-Quran manakala teknik menggunakan pustaka asas mencatat keputusan terendah. Hal ini demikian kerana teknik ini tidak diubah khas mengikut domain seperti dua teknik lain. Teknik hanya menggunakan entiti asas yang dibekalkan oleh SpaCy. Teknik berasaskan peraturan dan teknik pembelajaran mesin mempunyai dapatan semula yang lebih rendah berbanding teknik menggunakan pustaka asas kerana kedua-dua teknik ini gagal mengekstrak perkataan atau frasa yang boleh dilabel sebagai entiti. Kebanyakannya berpunca daripada entiti nama golongan. Hal ini demikian kerana terdapat perbezaan antara teks Al-Quran dalam bahasa Arab dan terjemahan bahasa Inggeris. Situasi ini

dapat dilihat melalui contoh ayat ke-5 surah Al-Qasas yang digunakan dalam penilaian.



Rajah 5.4: Ayat ke-5 surah Al-Qasas

*And We wished to do a favour to those who were weak in the land and to make them rulers and to make them **the inheritors***

Rajah 5.5: Teks terjemahan ayat ke-5 surah Al-Qasas

Dalam bahasa Arab, perkataan yang ditandakan dalam rajah 5.4 merupakan entiti nama golongan. Namun begitu, teks terjemahan yang digunakan tidak menggunakan perkataan 'Warithoon' sebaliknya menggunakan frasa 'the inheritors' seperti yang ditunjukkan dalam rajah 5.5. Oleh itu, kedua-dua teknik tidak dapat mengecam perkataan tersebut sebagai entiti. Situasi ini berlaku kepada banyak entiti dalam teks terjemahan seperti frasa 'the disbelievers' dan 'those who give sincere advice' yang boleh dikategorikan sebagai entiti nama golongan.

6 KESIMPULAN

Sebagai kesimpulan, projek ini telah berjaya mencapai objektif kajian iaitu untuk melakukan pengecaman entiti nama terhadap teks terjemahan Al-Quran, membandingkan dan memilih teknik yang terbaik dalam mengekstrak maklumat daripada Al-Quran menggunakan NER. Teknik terbaik yang diperolehi melalui penilaian yang telah dibuat ialah teknik pembelajaran mesin berarah. Melalui projek ini, kaedah pra-pemprosesan data bagi data dalam domain Al-Quran dapat dikenal pasti. Projek ini juga berjaya mengekstrak entiti yang dapat dijadikan sebagai bahan kajian dalam bidang ontologi Al-Quran. Projek ini juga menunjukkan keperluan dalam pembinaan keperluan khas bagi mengekstrak entiti daripada Al-Quran kerana teknik menggunakan pustaka SpaCy asas menunjukkan pencapaian yang sangat tidak

memuaskan jika dibandingkan dengan dua teknik yang dibina khas untuk domain Al-Quran. Namun begitu, penambahbaikan dalam pembinaan model pembelajaran mesin dan pembinaan peraturan dapat dilakukan dalam kajian pada masa hadapan bagi mengekstrak lebih banyak kategori entiti dengan ketepatan yang lebih tinggi.

7 RUJUKAN

- Cyril Goutte & Eric Gaussier. 2005. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. *Lecture Notes in Computer Science*. 3408. 345-359.
- Gabriel Reynolds. 2012. 'Two Unique Translations of the Quran'. International Quranic Studies Association. <https://iqsaweb.wordpress.com/tag/muhammad-muhsin-khan/>.
- Maha Al-yahya, Hend Al-Khalifa, Bahanshal, A., & Nawal Al-Helwah. 2010. An Ontological Model for Representing Semantic Lexicons: An Application on Time Nouns in the Holy Quran. *The Arabian Journal for Science and Engineering*. 35.
- Neelam Tyagi. 2020. What is spaCy in Natural Language Processing (NLP)? / *Analytics Steps*. AnalyticsSteps. <https://www.analyticssteps.com/blogs/what-spacy-natural-language-processing-nlp>
- Rohana Ismail, Zainab Abu Bakar, & Nurazzah Abdul Rahman. 2015. Extracting knowledge from english translated quran using NLP pattern. *Jurnal Teknologi UTM*, November.
- Wilfred Van Casteren. 2017. The Waterfall Model and the Agile Methodologies : A comparison by project characteristics. 10.13140/RG.2.2.36825.72805.